

A Forecasting Approach to Online Change Detection in Land Cover Time Series

Willem C. Olding[✉], Jan C. Olivier, Brian P. Salmon[✉], and Waldo Kleynhans

Abstract—We present a method for online detection of land cover change based on remotely sensed time series. Change is detected by monitoring deviations between observations and forecasts made using the time series historical data and similar time series in the geographical region. This method and several others were applied to MODIS 8-day surface reflectance data for problems of detecting settlement expansion in Limpopo Province, South Africa, and detecting deforestation in New South Wales, Australia. The proposed method had significantly shorter median detection delay (DD) for equivalent rates of false alarms compared with the other evaluated methods. We obtained a median DD of seven samples for settlement detection and 14 samples for deforestation detection corresponding to 56 days and 112 days, respectively. This is compared with a median DD of 224 and 544 days for the best other methods evaluated. We suggest that the proposed method is an excellent candidate for land cover change detection where rapid detection is essential.

Index Terms—Change detection algorithms, remote sensing, sequential detection, time series analysis.

I. INTRODUCTION

ONE of the major goals of remote sensing satellite systems is to enable large scale monitoring of the earth's land cover. Detecting when and where land cover change has occurred forms an important part of the monitoring process. Change detection is typically used in one of three cases. It may be used as the first stage in a processing pipeline, for example, when updating land cover maps, it can reduce the problem to only reclassifying the regions which have changed. It may also be applied to produce useful information in its own right, for example, estimating how much of a particular region has experienced change within a given time span. The third case, which will be the focus of this paper, involves monitoring land cover in near real time in order to detect change shortly after it has occurred. In this case, the output of a method can be used to guide action in response to the change.

Pixel time series generated by repeated observations from high temporal resolution multi-spectral satellite platforms have been shown to be well-suited to detecting changes in vegetated

land cover [1]–[9]. Detection of change from pixel reflectance time series is made difficult by several factors. Difficult to predict inter-annual variations coupled with semi-regular seasonal oscillations in reflectance make the time series highly non-stationary. This makes detecting land cover change a difficult problem requiring specialized methods [2]. Second, exactly what constitutes land cover change can be difficult to define. In this paper, the focus is on detecting human induced changes that take place over a 6–12 month period. This is typical of the types of land cover change associated with gradual clearing of forest or urban expansion.

One advantage of remotely sensed time series is the availability of many similar land cover time series within a region of interest. This collection of time series can be used to estimate the natural variation in reflectance caused by the influence of climatic changes on vegetation phenology. In this paper, we propose an online change detection method that used this technique to detect deviations of a time series from its natural variations implying a change in the underlying land cover.

Common themes in existing land cover time series classification methods involve transformation of the time series into a form that is approximately stationary under no-change conditions. Classical change detection methods that assume stationarity can then be applied [10]. Some approaches assume the original time series is cyclo-stationary [6] while others make assumptions regarding the structure of the trend and periodicity [3], [7], [8]. These types of methods can be further split into supervised methods, which require a labeled training set to estimate class specific statistics [6], [9], and unsupervised methods, which have no such requirement [3], [5], [7], [8].

In [6], change is detected from one known class to another by estimating class specific densities and calculating the likelihood ratio. The log-likelihood ratio was monitored for change using Page's cumulative sum (CUSUM) algorithm. To account for seasonal variation, a density was calculated for each class for each time step throughout the year. Different years were aggregated, so inter-annual variation was not explicitly accounted for.

A supervised approach is presented in [9]. They propose training a probabilistic classifier (random forest) at each time sample to estimate time series of class membership probabilities. These are monitored for change using the MODTrendr Algorithm [4]. As a supervised method, this requires a large number of labeled no-change time series to fit the probabilistic classifiers. It also results in a large number of time series to monitor for change if there are many classes being considered.

Manuscript received May 7, 2018; revised September 6, 2018 and February 9, 2019; accepted March 7, 2019. Date of publication May 5, 2019; date of current version May 25, 2019. (Corresponding author: Willem C. Olding.)

W. C. Olding, J. C. Olivier, and B. P. Salmon are with the School of Engineering, University of Tasmania, Hobart, TAS 7005, Australia (e-mail: Willem.Olding@utas.edu.au; JC.Olivier@utas.edu.au; Brian.Salmon@utas.edu.au).

W. Kleynhans is with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria 0081, South Africa (e-mail: wkleynhans@csir.co.za).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2905594

The work presented in [5] details a method designed specifically for detecting change in periodic, non-stationary time series, not limited to land cover. It operates by using Gaussian process regression to make a probabilistic forecast of the next sample. Once the next sample is available, its z -score under the forecast distribution is calculated. By repeatedly applying this forecasting method, a time series of z -scores is produced. Given the model is working correctly, the resulting time series of z -scores should be Gaussian independent identically distributed (IID). A series of samples that deviate significantly from the forecast will result in a change in distribution of the series of z -scores. This is detected using an exponentially weighted moving average control chart. The kernel for the Gaussian process was a periodic exponential allowing for modeling of periodic functions. The parameters for the kernel were fit at the start of the signal and then held constant. The computational performance and scalability of the method is improved upon in [11].

In [3], a parametric model comprised of a bias, linear trend, and periodic seasonal component is used to make predictions. Three periodic terms are included in the model, a fundamental frequency equal to one cycle per year and two additional harmonics. This model is extrapolated into the future and the moving sum (MOSUM) of residuals over a finite lookahead window is monitored for deviation from the fitted model. The probability of observing a particular MOSUM with respect to the time series noise level is derived and a change alarm is triggered when it exceeds some significance level.

A parametric model-based method is also applied in [7]. The authors take the approach of fitting a triply-modulated cosine and bias to land cover time series over a sliding window. The time series of parameters from each sliding window step is monitored for change. It is suggested that the bias parameter of the model should remain constant during a period of no-change and deviate when the land cover experiences change. The bias parameters series is converted to a series of z -scores by estimating parameters from its historical values. Change is declared when six out of ten previous z -scores exceed some threshold.

The work presented in [8] applies a parametric model identical to that of [3]. In this case, residuals from the model over the entire series are converted to z -scores and this time series is directly monitored for outliers. z -scores exceeding a given significance level are considered indicative of a change point. While this method uses a similar approach to those mentioned earlier, it cannot be considered as an online approach as it requires the entire dataset data in order to fit the parametric model and thus can only discover changes in the past.

The method we propose combines the ideas of the works given in [3] and [5] of using the prior samples of a time series to make a forecast with ideas of [6] and [9] of considering samples of other time series in the same region to estimate a baseline from which to detect change. Under our unified framework, a collection of similar time series from the same region are used to estimate the joint distribution of the time series over a temporal window. This is conditioned on previous observations of a signal in order to make a prediction and calculate z -scores. The time series of z -scores is then monitored for change.

II. BACKGROUND

A. Stationary of Land Cover Time Series

A stochastic process can be said to be *strictly stationary* if the joint distribution of any combination of values is time invariant. For a strictly stationary process, $\{X_t\}$, we can say for any set of time indices $\{t_1, \dots, t_k\}$, the joint distribution

$$\Pr(X_{t_1+\tau}, \dots, X_{t_k+\tau}) \quad (1)$$

is identical regardless of the time shift τ . A special case of stationary is IID. As the name suggests each sample is independently drawn from the same distribution. In this case, the joint distribution may be factorized as

$$\Pr(X_{t_1+\tau}, \dots, X_{t_k+\tau}) = \prod_{i=1}^k \Pr(X_{t_i+\tau}). \quad (2)$$

A process is said to be *cyclo-stationary* if the process obtained by sampling at integer multiples of some period T exhibits stationarity. Formally

$$\Pr(X_{t_1+T+\tau}, \dots, X_{t_1+nT+\tau}), n \in \mathbb{Z} \quad (3)$$

remains identical for any value of t_1 and τ and for some period T .

Several factors combine to make remotely sensed land cover time series fail to meet these stationarity assumptions. Most land cover is vegetated and this results in the dominant source of non-stationarity in the signal. It is well documented that the reflectance of vegetation varies approximately periodically in a yearly cycle [1]. In many cases, this may result in cyclo-stationarity however the magnitude and bias of the oscillations is often non-constant between years due to inter-annual climatic changes such as droughts. In many cases, trend components are present in vegetated time series that suggest either long term non-periodic change or change on a cycle much longer than the time span of available data.

III. CHANGE DETECTION BY FORECASTING

Let us consider an estimator that is able to produce forecasts of the next value in a time series under no-change conditions. We denote the forecast at time t as \hat{x}_t . The residual of the forecast is defined as $r_t = x_t - \hat{x}_t$. A sufficiently good estimator should yield a series of residuals, which are uncorrelated, unbiased (i.e., zero mean), and Gaussian. If the estimated residuals also have constant variance we can say the resulting time series of residuals is IID. Furthermore, if the residuals do not have constant variance but the estimator is able to produce good estimates of the variance of each sample, σ_t^2 , the residuals can be converted to z -scores by

$$z_t = \frac{r_t}{\sigma_t} \quad (4)$$

and the series of z -scores will be IID unit Gaussian.

If the original time series contains a change point that is unforeseen to the estimator, this will induce a bias in the series of z -scores/residuals, which can be detected by a classical change detection method. An ideal estimator should yield a series of

z-scores that is Gaussian IID under normal circumstances and respond strongly and rapidly in the presence of a change point. Such an estimator may take advantage of historical data, neighboring pixel time series, or ancillary data in order to make its prediction.

One such method that follows this general framework is the GPChange method of [5] and [11]. In this case, the prediction is made using only historical data from the same signal

$$\Pr(X_t) = \Pr(X_t | X_{t-1}, X_{t-2}, \dots, X_0). \quad (5)$$

Motivated by these previous works, we propose a method that takes advantage of both the previous samples in the time series and also the time series in the surrounding region when making a prediction. This is done by directly estimating the joint distribution over a temporal window using a set of similar time series. As the model is re-estimated at each time step this does not require any assumptions of stationarity. Furthermore it does not impose assumptions on the structure of the data as in [3], [7], and [8].

A. Proposed Method

We propose modeling the joint distribution over the samples in a temporal window as a multivariate Gaussian. Such a model can capture the mean signal of the region of interest, the uncertainty associated with the mean signal and the correlations between samples in the time series. The resulting model can be considered in a similar sense to a Gaussian process although discrete and with covariance estimated directly from data rather than predefined by a covariance function. By estimating the joint distribution from nearby time series, it is possible to capture the natural variations of a region. This statistical information can be leveraged to detect when a time series moves in a way that is inconsistent with both its own historical trajectory and that of the surrounding region.

With access to a fitted joint distribution, it is possible to condition on the previous observations in a signal to obtain a univariate distribution over the last sample. More explicitly, let $\mathbf{x} = [x_t, \dots, x_{t-W-1}]^T$ be a vector of time series observations in a temporal window of size W . Let $\mathcal{X} = \mathbf{x}^{(i)}_{i=1}^{i=D}$ be a set of D time series from a similar geographical region and ideally containing similar land cover. We can assume that each of these time series windows is an observation of a multivariate random variable $\mathbf{X} = [X_t, \dots, X_{t-W-1}]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Under the multivariate Gaussian assumption, we can calculate the mean vector and covariance matrix as

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] \quad (6)$$

and

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \quad (7)$$

which can be estimated directly from the data.

For a single time series, $\mathbf{x}^{(i)}$, this model can be conditioned to calculate the univariate distribution over the last time sample given earlier observations within some temporal window, $\Pr(x_t^{(i)} | \mathbf{x}_{\text{obs}}^{(i)})$, using the standard method for conditioning a multivariate Gaussian. If we partition the covariance matrix into

observed and unobserved variables as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_t^2 & \boldsymbol{\sigma}_{1,2} \\ \boldsymbol{\sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix} \quad (8)$$

where

$$\boldsymbol{\sigma}_{1,2} = \boldsymbol{\sigma}_{2,1}^T = [\text{cov}(X_t, X_{t-1}), \dots, \text{cov}(X_t, X_{t-W-1})] \quad (9)$$

and

$$\boldsymbol{\Sigma}_{2,2} = \begin{bmatrix} \sigma_{t-1}^2 & \dots & \text{cov}(X_{t-1}, X_{t-W-1}) \\ \vdots & \ddots & \vdots \\ \text{cov}(X_{t-W-1}, X_{t-1}) & \dots & \sigma_{t-W-1}^2 \end{bmatrix} \quad (10)$$

the conditional distribution is univariate Gaussian with mean and variance given by

$$\bar{\mu}_t^{(i)} = \mu_t + \boldsymbol{\sigma}_{1,2} \boldsymbol{\Sigma}_{2,2}^{-1} (\mathbf{x}_{\text{obs}}^{(i)} - \mu_{\text{obs}}) \quad (11)$$

and

$$\bar{\sigma}_t^2 = \sigma_t^2 - \boldsymbol{\sigma}_{1,2} \boldsymbol{\Sigma}_{2,2}^{-1} \boldsymbol{\sigma}_{2,1}. \quad (12)$$

From this probabilistic forecast, the z-score of the observation of the i th time series can be calculated as

$$z_t^{(i)} = \frac{x_t^{(i)} - \bar{\mu}_t^{(i)}}{\bar{\sigma}_t^{(i)}}. \quad (13)$$

Repeatedly applying this method as the sliding window is shifted results in a time series of z-scores. The time series of z-scores can then be monitored for persistent deviation from the standard normal, which implies a change.

If the region contains a very diverse land cover classes, it might be the case that the variance of the fitted multivariate Gaussian is too high to produce useful forecasts. We suggest two alternative methods for estimating the joint Gaussian density. The choice of method depends on domain knowledge of the land cover types in the region of interest. In the case that there is one single dominant land cover type that should be monitored for change, the robust estimator of [12] is well suited. This method estimates the mean and covariance matrix using a fixed size subset of the samples. The subsample is selected from all possible combinations as the one that yields the covariance matrix with the smallest determinant. This is very effective at eliminating outlying time series in the dataset.

Another possibility is that there are multiple types of land cover type present in the region of interest. In this case, a Gaussian mixture model approach can be used to simultaneously estimate multiple Gaussian joint densities and group similar time series. The joint distribution of the set of time series is then expressed as

$$\Pr(\mathbf{X}) = \sum_{c=1}^k w_c \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c). \quad (14)$$

The weights, w_i , and parameters of the Gaussian distributions are estimated using the expectation maximization algorithm. The number of components must be selected in advance using location specific knowledge. When making a forecast, the cluster that a time series belongs to can be estimated by finding

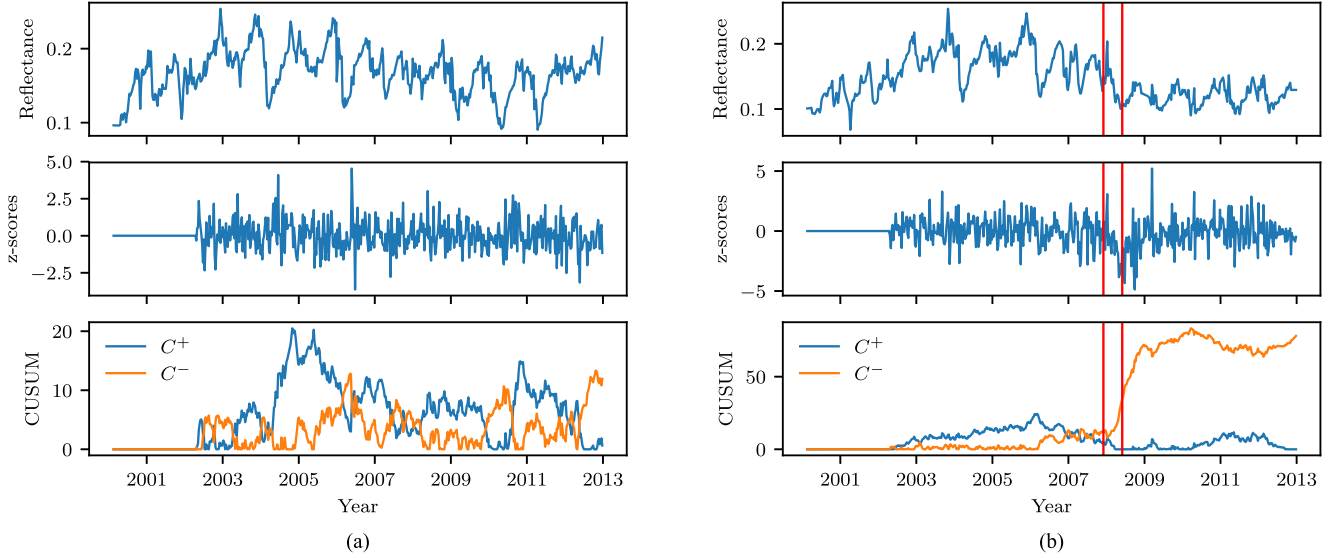


Fig. 1. Reflectance time series, z-score series, and CUSUM output for a no-change series and a synthetic change from the Limpopo dataset. Vertical bars indicate the beginning and end of the linear transition. (a) No change time series. (b) Change time series.

closest component mean using the samples in the window. The mean vector and covariance matrix from this component are then used in (11) and (12). This is the variation we focus on in this study.

B. Change Detection on z-Scores

Once the time series of z-scores has been produced by one of the methods discussed earlier it can be monitored to detect deviations from its expected distribution. Provided there are no changes present in the time series and the forecasting algorithm is working correctly, the z-score time series should be IID Gaussian. To detect statistically significant persistent deviations from this distribution, we apply Page's CUSUM algorithm [10], [13].

We present the CUSUM method under the assumption that the data are IID Gaussian with known mean and variance prior to change, after which it experiences a change in mean of unknown magnitude.

Let $\{z_i\}_{i=0}^{i=N}$ be the series of z-scores. By definition, this series should be IID Gaussian with zero mean and unit variance. We therefore expect under normal conditions that the CUSUM

$$C_i = \sum_{j=1}^i z_j \quad (15)$$

$$= z_i + C_{i-1} \quad (16)$$

is a random walk with zero mean [10]. If there is a persistent bias in the models ability to make predictions, such as the underlying data generating distribution changing, we expect the random walk to drift away from zero in the direction of the bias. To improve the robustness of the method, a slack parameter, k , is included [10]. This parameter is subtracted from the CUSUM at each iteration and reduces the sensitivity to small mean shifts, which can help prevent false alarms. The CUSUMs are also split into positive and negative components with each clamped

at zero. The resulting CUSUMs are defined recursively as [10]

$$C_i^+ = \max[0, z_i - k + C_{i-1}^+] \quad (17)$$

$$C_i^- = \max[0, -z_i - k + C_{i-1}^-]. \quad (18)$$

As we do not know how the predictor will respond to a land cover change, we monitor both CUSUMs. When either C^+ or C^- exceeds some threshold λ a change alarm is triggered and the corresponding CUSUM is reset to zero. Fig. 1 shows the reflectance time series, z-score series extracted using the joint estimation method and the corresponding CUSUMs. In the time series containing a change, it is possible to observe a negative bias in the z-score series immediately after the change point. This causes a strong response in C^- . In the no-change series, the corresponding z-score series should be close to IID Gaussian noise and the CUSUMs of this signal will not significantly deviate from zero.

It can be shown that if the mean and variance before and after the change point is known and the change takes place instantaneously, this method is optimal for a given detection delay (DD) [14]. This does not hold in our case as the distribution of the z-scores post change cannot be known.

IV. STUDY AREAS AND DATA

A. Study Areas

We evaluate the methods by applying them to two applications that require rapid and accurate change detection.

The first problem is that of detecting settlement expansion in the Limpopo Province of South Africa. In some areas of this province settlement, expansion is carried out *ad hoc* by residents without prior planning approval. The local native vegetation of this region is predominately Savannah, characterized by small sparse trees and grassland [15]. Construction of settlements is

characterized by the clearing of native vegetation and the construction of small dwellings [16]. Detection of unplanned settlements in this region has been the focus of several studies [6], [17]–[19].

The second problem relates to detecting changes in areas of native forest in New South Wales (NSW), Australia. Under the Australian government emissions reduction fund initiative, land owners are able to generate Australian carbon credit units in exchange for preventing deforestation of native forest for which a clearing permit has previously been issued. To be eligible, the native forest must have at least 20% canopy coverage with tree height greater than two meters [20].

At the commencement of a project, an extensive audit is undertaken by an accredited third party and estimates made of the appropriate number of credits to be allocated throughout the project. A number of subsequent audits are also required to ensure compliance throughout the project lifetime. Projects have a permanence period of either 25 or 100 years. During this time period, no clearing, with the exception of minimal thinning (<5%), is permitted [20].

At the time of writing, there are approximately 400 vegetation projects underway in Australia. A system of automated change alarms based on remotely sensed time series has the potential to significantly reduce the auditing workload and target it to locations where the forest cover has changed and the project may require reassessment. Continuous monitoring is also important to guarantee the integrity of the carbon credit units. We apply online change detection methods to several regions in rural NSW which are currently generating carbon credits in exchange for deforestation prevention.

B. Data Preparation

For both applications, we make use of time series from the MODIS MCD43A4 product [21]. This product delivers bi-directional reflectance distribution function corrected surface reflectance data from both the Terra and Aqua Platforms. The temporal resolution is eight days, with each data point selected as the highest quality acquisition in a 16-day temporal window. The spatial resolution is approximately 500 m. This product has been used in numerous land cover change detection studies mostly due to its high temporal resolution and large catalog of data.

The dataset for the Limpopo region is comprised of $D = 2348$ MODIS pixel time series over the time span of the years 2000 to 2013. Each was classified by an expert analyst as *vegetation* - 997 (42%), *settlement* (53%) - 1235, and *change* - 116 (5%) by visual inspection of high resolution imagery. A pixel was labeled as change if it transitioned from vegetation to settlement within the time span of interest. The exact time of change is not known. A comprehensive description of the dataset can be found in [19]. This type of change is classified as a land cover conversion. To quantify the DD of the methods in this study, it was necessary to create synthetic change time series where the exact time of change can be controlled. To generate synthetic change time series, we follow a similar approach to [18]. Two time series of different land cover types are linearly blended over

a period of six months. This slow transition between land cover types aims to imitate the gradual conversion that is expected in the construction of a settlement.

The dataset for NSW is comprised of $D = 1994$ MODIS pixel time series located within regions assigned to a project which is currently earning carbon credits in exchange for avoided deforestation. These time series span the years 2008 to 2018. Each time series has a corresponding pair, which is closely located and contains a similar type of land cover but at a lower density. As no real change was observed within the designated time span, only synthesized change series were considered. These change series were created by linearly blending over a period of six months between a pixel and its lower density pair. This aims to simulate the partial and gradual clearing of vegetation. Fig. 3 shows the study region and an example of two nearby pixels of differing densities. This problem is an example of within-class change detection [1].

The remotely sensed time series for both of these locations contained only a few missing values (<0.1%). The missing values present were filled using linear interpolation.

We consider only MODIS band 1 (620–670 nm) and band 2 (841–876 nm). These have been shown in previous studies [6], [19] to be most effective at detecting changes in vegetation. We also consider the normalized difference vegetation index, a commonly used quantity defined for MODIS as the difference between bands 2 and 1 divided by their sum.

V. EXPERIMENTAL METHODOLOGY AND RESULTS

A. Evaluated Methods

We compare three methods for extracting time series of z -scores that are ideally IID under no-change conditions: our proposed method of directly estimating the joint distribution using the Gaussian mixture model (Joint), a supervised univariate density estimation method similar to that of [6] (univariate) and a parametric forecasting method similar to that of [3] (parametric).

The supervised univariate approach assumes knowledge of the land cover class of all time series in the training set as well as the initial class prior to change. We fit a univariate Gaussian density for each class at each time step. At test time, z -scores are calculated using the appropriate density at each time step assuming the class does not change from its initial assignment.

In the parametric case, we fit a harmonic model with a bias, seasonal oscillation component, and two higher harmonics over a fixed length look-back window. The model is refit at each time step as the window is shifted. The parametric model is used to make a forecast and the difference between observation and prediction is converted to a z -score by dividing by the standard deviation estimated from the model residuals over the window.

Fig. 2 shows a comparison of the histogram and autocorrelation of the z -score series. This was extracted by each method for a single change and no-change time series from the Limpopo dataset. Recall that under no change conditions an ideal z -score series should be uncorrelated and Gaussian distributed. The estimated histogram and mean/variance gives an indication of

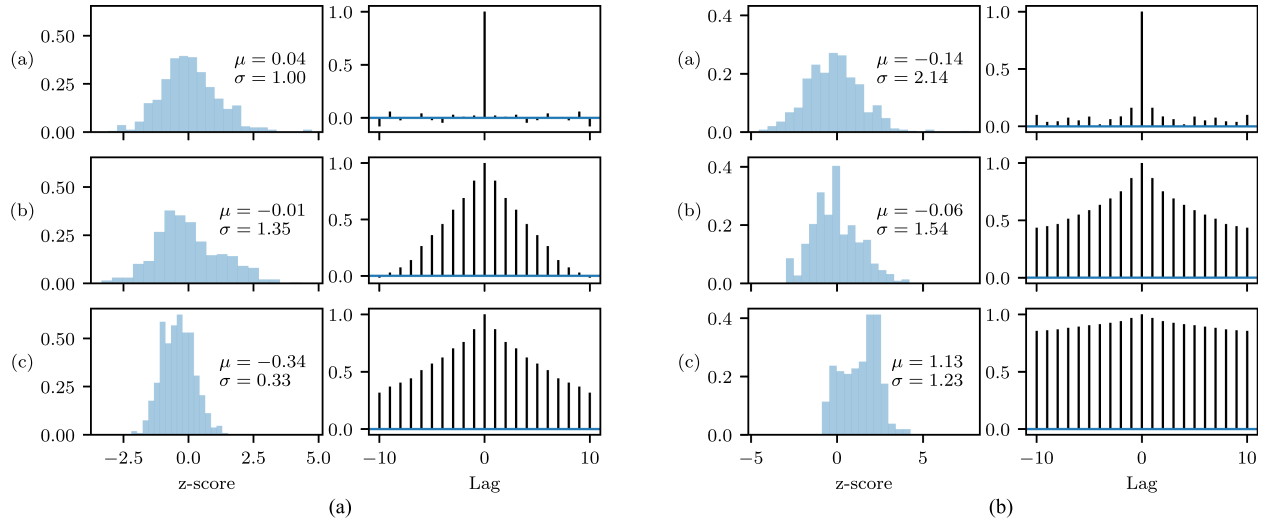


Fig. 2. Density plot and autocorrelation plot for each of the z-score extraction methods investigated in this paper. (a) Joint (GMM Method). (b) Parametric. (c) Univariate supervised. Evaluated on a no-change time series and a synthetic change time series from the Limpopo dataset using MODIS band 1. (a) No change time series. (b) Change time series.

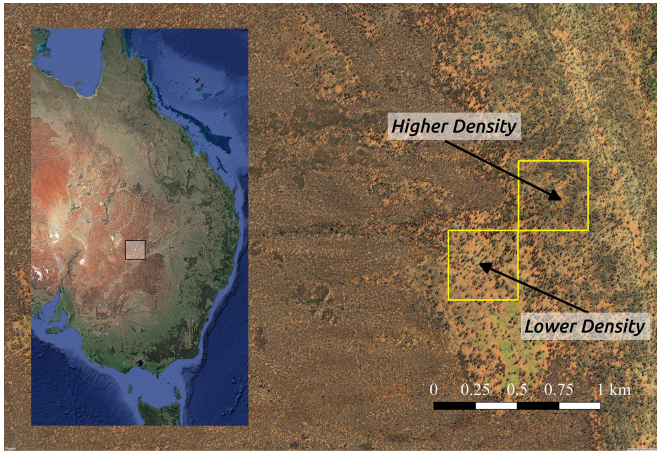


Fig. 3. Study region in New South Wales. This includes an example of two MODIS pixels within a protected area that contain similar land cover with differing densities. This pair of pixel time series is blended together over a six month transition to simulate the partial and gradual clearing of vegetation. (Image courtesy of Google Earth, DigitalGlobe.)

the distribution of the scores. To verify if the scores are independent, we can observe the estimated autocorrelation function. A perfectly uncorrelated/independent series should have an autocorrelation that is an impulse centered at zero lag. From inspection of the plots for a no-change time series the autocorrelation plot of the joint estimation methods suggest that the series of z-scores is very close to IID Gaussian. Both the parametric and univariate methods appear to have a much higher degree of correlation between samples in the z-score series. This is expected for the parametric model approach as the small number of harmonics limits the ability of the model to follow fast changes leading to correlated errors. All methods show an increase in correlation in a time series containing land cover change.

TABLE I
RESULTS FOR DETECTION OF SETTLEMENT EXPANSION IN
LIMPOPO PROVINCE USING MODIS BAND 1

Method	AUROC	TPR at FPR=0.2	DD at RLFA=200
GMM Joint	0.77	0.78 (SD=0.08)	7
Parametric	0.79	0.71 (SD=0.08)	38
Supervised	0.90	0.86 (SD=0.03)	28

TABLE II
RESULTS FOR DETECTION OF VEGETATION THINNING IN
NEW SOUTH WALES USING MODIS BAND 1

Method	AUROC	TPR at FPR=0.2	DD at RLFA=200
GMM Joint	0.81	0.68 (SD=0.08)	14
Parametric	0.72	0.57 (SD=0.12)	68
Supervised	0.65	0.42 (SD=0.06)	inf

B. Change Detection Assessment

To assess the change detection performance of the methods, we make use of the receiver operator characteristic (ROC) curve, which plots false alarm rate against true positive rate. This allows us to compare methods irrespective of a choice of threshold. Furthermore, it does not require us to arbitrarily select an acceptable false alarm rate, which may be application specific. We also tabulate the area under ROC curve (AUROC) and true alarm rate at a false positive rate of 0.2 in Tables I and II. An ROC curve gives an estimate of the correct classification rates, irrespective of delay, when applying the methods to time series of fixed length and assuming at most one alarm is triggered per series.

All methods requires selection of the slack parameter, k , and both the parametric and the joint Gaussian estimation methods require a finite length look-back window of size W in order to make a forecast. A larger window permits using more historical

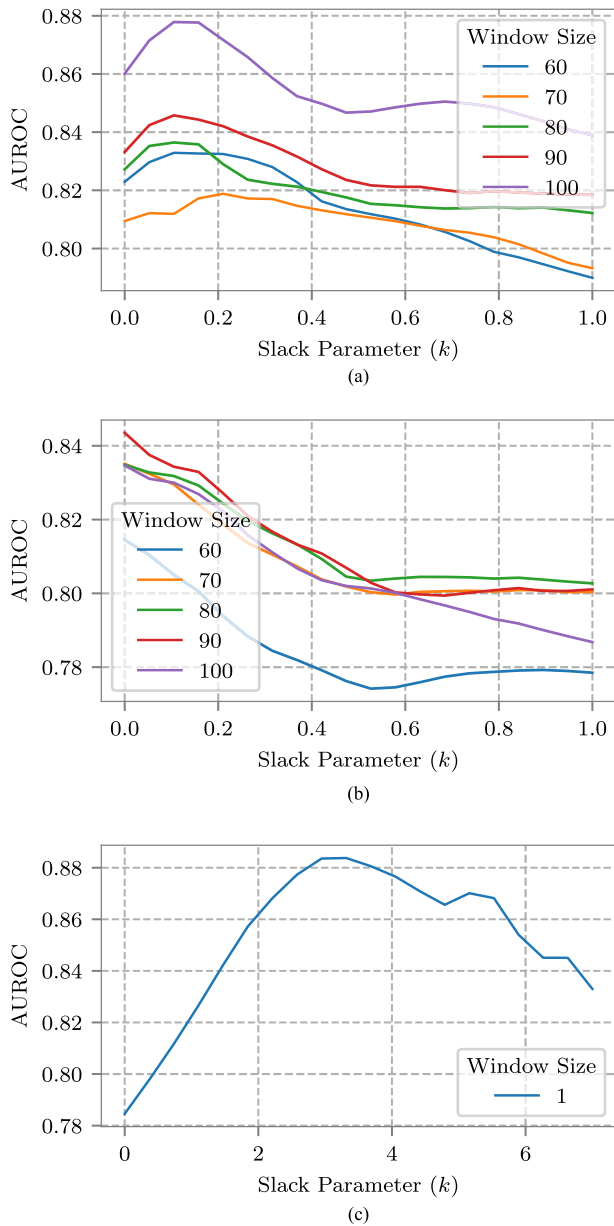


Fig. 4. AUROC as a function of CUSUM slack parameter, k . Plotted for multiple selections of window size for the Limpopo dataset, MODIS band 1. (a) Joint Gaussian mixture model method. (b) Parametric method. (c) Supervised univariate method.

data at the expense of computational complexity. This is especially true in the joint Gaussian case where a $W \times W$ covariance matrix must be stored and conditioned upon.

To visualize the influence of these parameters, Fig. 4 gives the AUROC for detection of change/no-change time series as a function of k for several window sizes. For each of the methods, it appears that a larger look-back window generally results in improved detection performance. We did not evaluate window size greater than 100 samples as this prevents the detection of changes occurring near the beginning of the series. In a production system, this could be increased or even allowed to grow as more data becomes available. The Gaussian joint density estimation method achieved the highest AUROC on the real

dataset with a slack parameter setting of around $k = 0.1$. It also appears insensitive to slack parameter selection within the evaluated range. The parametric z-score method was found to work best with a slack of zero and the supervised method with a slack of 3.0. Recall that the slack parameter is related to the expected change in mean that the z-score series should experience under land cover conversion. These results suggest that the sensitivity of the parametric method to change is already low while the univariate method has a high sensitivity that results in excessive false alarms without sufficient slack.

The ROC curve was evaluated for a window size of 100 and with the best slack parameter selected from the curves in Fig. 4. Fivefold cross validation was used to estimate the mean ROC curve and its 95% confidence interval. For the Limpopo dataset, real change data was used to generate while for the NSW dataset synthetic change series were combined with real no-change time series. Fig. 5(a) gives the curves for the Limpopo dataset and Fig. 6(a) gives the curves for the NSW dataset. It can be seen that when considering detection on the Limpopo dataset there is no significant difference between each of the assessed methods with the supervised estimation method giving the highest AUROC. On the NSW dataset, the supervised method did not perform well showing significantly inferior performance to the other methods. This is likely due to the poorly defined land cover classes (“dense” and “sparse”) in this dataset. In this case, GMM joint estimation method showed the best performance. Comparison between the Limpopo and NSW datasets illustrates the increase in difficulty in detection within class change compared with between class change.

C. DD Assessment

True/false positive rates give the probability of correct/incorrect detection at any time within the length of the signal. For continuous online detection, this does not give any indication on the expected delay until detection. Instead, we prefer to compare methods in terms of expected run lengths (i.e., time between alarms). This can be further split into two quantities, the run length to false alarm (RLFA) and the run length to true alarm or DD. The RLFA is defined as the number of samples after the algorithm starts, either at the beginning of the signal or after resetting due to a previous alarm, to the next alarm. The DD is defined as the number of samples between the actual change point and the next alarm. A good detection method strives to maximize the expected RLFA while minimizing the expected DD [10]. These can be considered as the generalization of false alarm rate and true alarm rate to the continuous monitoring case.

Estimating the distributions of RLFA and DD for a given method using finite length signals presents some challenges. As the z-scores signals are noisy, the CUSUM method will always trigger an alarm if it is allowed to run for a sufficiently long time [10]. When estimating run length distributions from finite length signals, this duration may be longer than the length of the signal. This can be taken into account by using the concept of censoring from survival analysis. The observation of no alarm gives information that the DD was at least as long as the remainder of the signal, known as right censoring. Failure to take this into

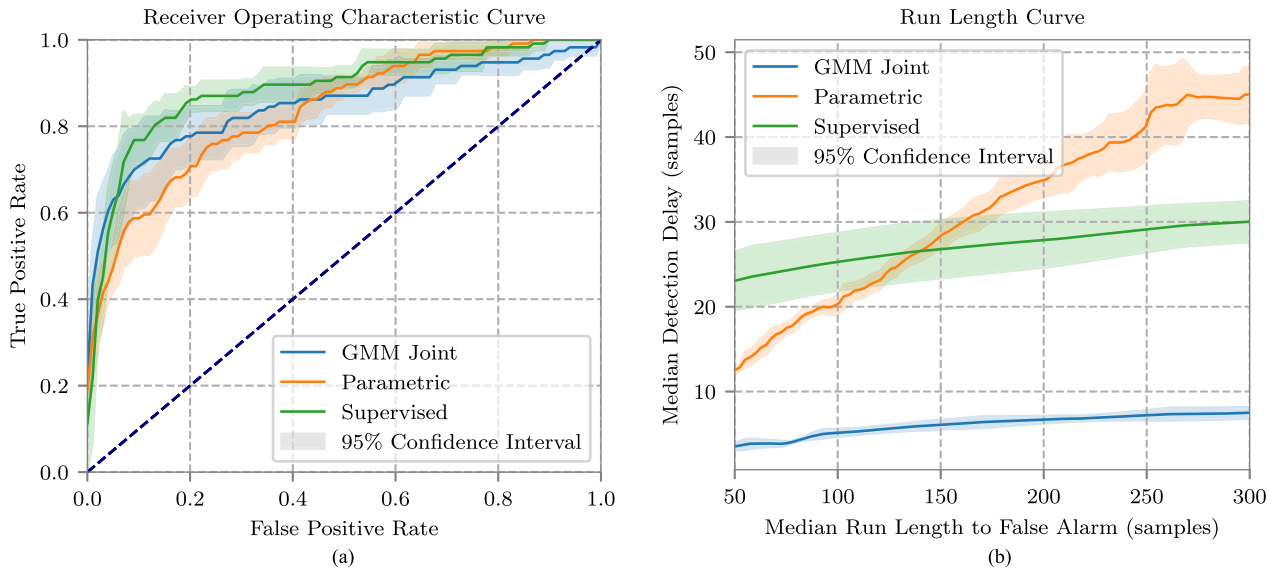


Fig. 5. ROC curves and DD curves for each of the considered methods applied to the Limpopo dataset using MODIS band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter determined from curves. The ROC curves were generated from real change time series and the DD curves from a combination of real no-change and synthetic change time series. (a) ROC Curves. (b) DD versus RLFA curves.

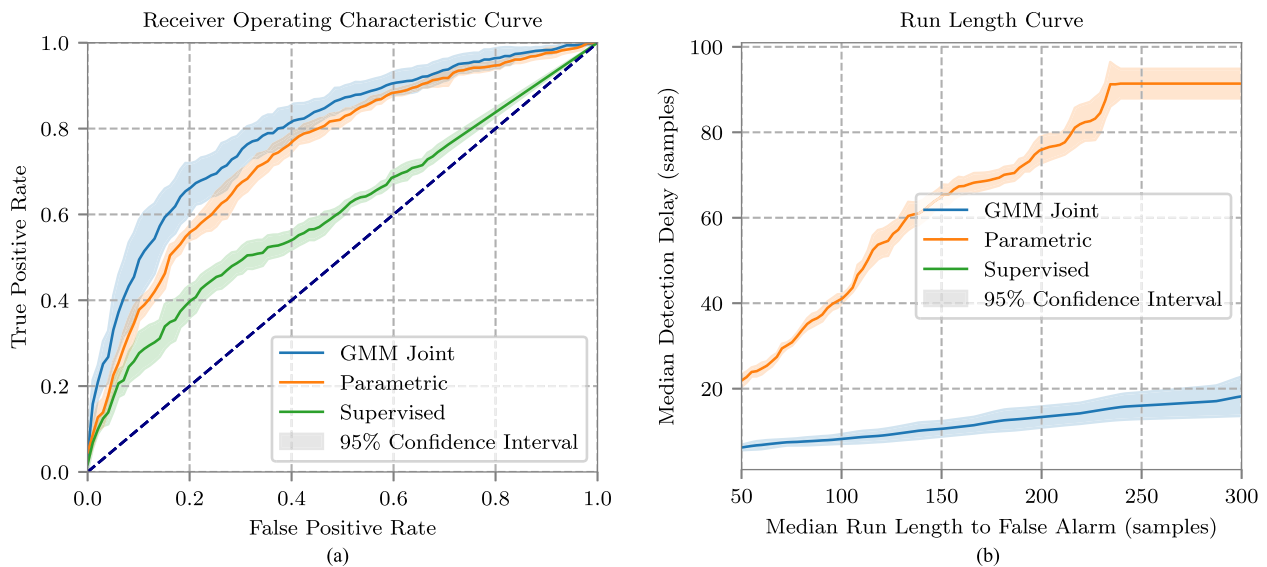


Fig. 6. ROC curves and DD curves for each of the considered methods applied to the NSW data using MODIS band 1. Curves were evaluated with a look-back window size of $W = 100$ and best slack parameter determined from curves. Both plots were generating using a combination of real no-change and synthetic change time series. (a) ROC curves. (b) DD versus RLFA curves.

account, for example, by dropping series with no alarm, results in an underestimate of run lengths. To account for the censoring of data, we estimate the median DD by fitting a Kaplan–Meier estimator [22], a commonly used non-parametric method for estimating a survival function in the presence of censored data.

A survival function was estimated for both the RLFA and DD for each method at a series of thresholds. For RLFA, if no alarm was triggered, the run length was said to be at least the length of the signal. The RLFA was also considered to be censored when an actual change point occurred. For DD, if no alarm was triggered after the change point, the run length was said to be at

least the difference between the change point and the end of the signal. Fig. 7 compares the cumulative distribution functions (CDF) (1 - survival function) for each method with a single threshold selected to yield a median RLFA of 200. The survival curves give much better insight than the median alone. It can be seen, for example, that 90% of changes in the Limpopo dataset and 70% in the NSW dataset are detected with a delay less than 30 samples (120 days) using the GMM joint approach.

Plotting the median DD versus median RLFA with changing threshold can be considered as a continuous inspection equivalent to an ROC curve. Fig. 5(b) gives the curves for the Limpopo

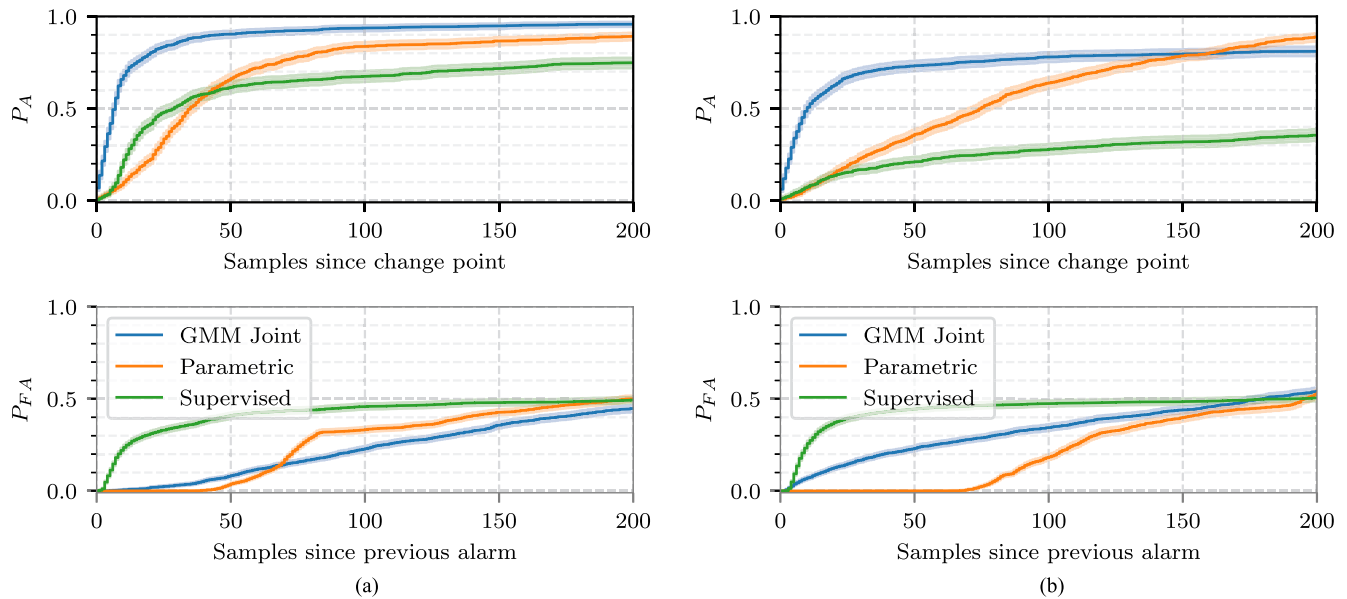


Fig. 7. Comparison of CDFs for detectors with a fixed thresholds. This gives the probability that an alarm (P_A) or false alarm (P_{FA}) has occurred after a given number of samples. The median delay or run length (plotted in Figs. 5(b) and 6(b) with varying threshold) is the point where the CDF has probability equal to 0.5. Thresholds were selected to yield a median RLFA of 200 samples. (a) CDFs for Limpopo dataset. (b) CDFs for NSW dataset.

dataset and Fig. 6(b) gives the curves for the NSW dataset. Again, the mean and 95% confidence intervals for each curve were estimated from fivefold cross validation. From both of these plots, it is clear that the GMM joint density estimation methods show a significant advantage in run length when compared with the other methods. Tables I and II give the median DDs for each method if the threshold is selected to yield a median RLFA 200 samples. For detecting settlement expansion, this illustrates that the median DD for the joint estimation method is four times shorter using the supervised univariate method and around five times shorter than the parametric method. With the MODIS sample period of eight days, this corresponds to a median wait of 56 days for the joint method, 224 days for the supervised method, and 304 days using the parametric method. This is a significant improvement on supervised CUSUM method of [6], which achieved a DD of approximately 400 days applied to synthetic settlement expansion data from the same region. For detecting deforestation in NSW, the DDs were slightly longer in comparison. However, the joint estimation method still outperformed the parametric method with a median DD almost five times shorter than the parametric method. The supervised method did not successfully detect enough change points to estimate the median using the Kaplan–Meier method and therefore, its curve could not be included. These delays are in the range of those reported in [7] for the detection of deforestation; however, it is difficult to compare as this paper does not report the accompanying RLFA.

VI. CONCLUSION

The forecasting method proposed in this paper was shown to give similar performance to two other online methods in terms of detection of a change point within a fixed length signal. The ROC

curves of Figs. 5(a) and 6(a) suggest that higher false alarm rates must be accepted to yield equivalent true positive rates to other offline methods in the literature applied to the same or similar datasets [17]–[19], [23]. This is a tradeoff that must be made for fast, online detection. We show the limitations of the ROC plot for quantifying performance in the continuous inspection case and suggest an alternative of plotting the estimated median run lengths to false alarm and DD.

When considering the continuous inspection case, the proposed method was shown to exhibit significantly shorter median DDs for a given median RLFA when compared with the other considered methods. These results suggest that the proposed method is a good candidate for land cover change detection system where rapid change detection is important. Future research should consider the generalization of the method to multivariate data to allow multiple bands to be used simultaneously, which may be able to further reduce the DD and occurrence of false alarms.

REFERENCES

- [1] R. S. Lunetta and C. D. Elvidge, *Remote Sensing Change Detection*. New York, NY, USA: Taylor & Francis, 1999.
- [2] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, “Phenological change detection while accounting for abrupt and gradual trends in satellite image time series,” *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2970–2980, 2010.
- [3] J. Verbesselt, A. Zeileis, and M. Herold, “Near real-time disturbance detection using satellite image time series,” *Remote Sens. Environ.*, vol. 123, pp. 98–108, 2012.
- [4] R. E. Kennedy, Z. Yang, and W. B. Cohen, “Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendrTemporal segmentation algorithms,” *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010.
- [5] V. Chandola and R. R. Vatsavai, “A Gaussian process based online change detection algorithm for monitoring periodic time series,” in *Proc. Int. Conf. Data Mining*. SIAM, 2011, pp. 95–106.

- [6] T. L. Grobler, E. R. Ackermann, A. J. van Zyl, J. C. Olivier, W. Kleynhans, and B. P. Salmon, "Using page's cumulative sum test on MODIS time series to detect land-cover changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 332–336, Mar. 2013.
- [7] A. Anees and J. Aryal, "Near-real time detection of beetle infestation in pine forests using MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3713–3723, Sep. 2014.
- [8] Z. G. Zhou and P. Tang, "Continuous anomaly detection in satellite image time series based on z-scores of season-trend model residuals," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 3410–3413.
- [9] H. Yin, D. Pflugmacher, R. E. Kennedy, D. Sulla-Menashe, and P. Hostert, "Mapping annual land use and land cover changes using MODIS time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3421–3427, Aug. 2014.
- [10] D. C. Montgomery, *Introduction to Statistical Quality Control*. New York, NY, USA: Wiley, 2009.
- [11] V. Chandola and R. R. Vatsavai, "A scalable Gaussian process analysis algorithm for biomass monitoring," *Statist. Anal. Data Mining, ASA Data Sci. J.*, vol. 4, no. 4, pp. 430–445, 2011.
- [12] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
- [13] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [14] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *Ann. Statist.*, vol. 18, pp. 1464–1469, 1990.
- [15] K. C. Steenkamp, K. J. Wessels, S. Archibald, and G. V. Maltitz, "Long-term phenology and variability of southern African vegetation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. 816–819. [Online]. Available: <https://doi.org/10.1109/IGARSS.2008.4779474>
- [16] P. Harrison, "Urbanization: The policies and politics of informal settlement in south Africa: A historical perspective," *Africa Insight*, vol. 22, no. 1, pp. 14–22, 1992.
- [17] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. van den Bergh, K. J. Wessels, and T. Grobler, "Detecting land cover change using a sliding window temporal autocorrelation approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 6765–6768.
- [18] W. Kleynhans, J. C. Olivier, K. J. Wessels, B. P. Salmon, F. Van den Bergh, and K. Steenkamp, "Detecting land cover change using an extended Kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 507–511, May 2011.
- [19] B. P. Salmon, D. Holloway, W. Kleynhans, J. C. Olivier, and K. J. Wessels, "Applying model parameters as a driving force to a deterministic nonlinear system to detect land cover change," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7165–7176, Dec. 2017.
- [20] "Carbon farming initiative avoided deforestation 1.1—Methodology determination," Australian Federal Register of Legislation, Mar. 2015.
- [21] C. Schaaf and Z. Wang, "MCD43A4 MODIS/Terra+Aqua BRDF/albedo nadir BRDF adjusted L3 global - 500m v006," NASA EOSDIS Land Processes DAAC, Sioux Falls, SD, USA, 2015. [Online]. Available: <https://lpdaac.usgs.gov/products/mcd43a4v006/#citation>
- [22] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.
- [23] B. P. Salmon, J. C. Olivier, W. Kleynhans, and K. J. Wessels, "Using the butterfly effect in a deterministic non-linear system to detect land cover change," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 4224–4227.



Willem C. Olding received the B.Eng. degree in computer systems engineering in 2013 from the University of Tasmania, Hobart, TAS, Australia, where he is currently working toward the Ph.D. degree.

His research interests include time series analysis, machine learning, signal processing, probabilistic graphical models, image analysis, and remote sensing.



Jan C. Olivier received the Ph.D. degree in electrical engineering from the University of Pretoria, Pretoria, South Africa, in 1990.

He is currently a Professor in communications engineering with the School of Engineering and ICT, University of Tasmania, Hobart, TAS, Australia. He was with Bell Northern Research, Ottawa, ON, Canada, and with Nokia Research Center in the U.S. His research interests include estimation and detection theory, as well as applications of machine learning.

Dr. Olivier is currently an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



Brian P. Salmon received the M.Eng. and Ph.D. degrees in electronic engineering from the University of Pretoria, Pretoria, South Africa, in 2008 and 2012, respectively.

He is currently a Post-Doctorate Research Fellow with the School of Engineering and ICT, University of Tasmania, Hobart, TAS, Australia, and also with the Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria, South Africa. His research interests include information theory, coding theory, and machine learning and graph theory.

Dr. Salmon was the lead author of the paper that was the recipient of the IEEE Geoscience and Remote Sensing Society 2012 Symposium Prize Paper Award.



Waldo Kleynhans received the Ph.D. degree in electronic engineering from the University of Pretoria, Pretoria, South Africa, and the MBA degree from Edinburgh Business School, Heriot-Watt University, Edinburgh, U.K.

He currently holds adjunct faculty positions with universities and science councils in South Africa as well as the USA. His research interests include the application of artificial intelligence and statistical signal (and image) processing in various fields, most notably, satellite remote sensing and telecommunications.